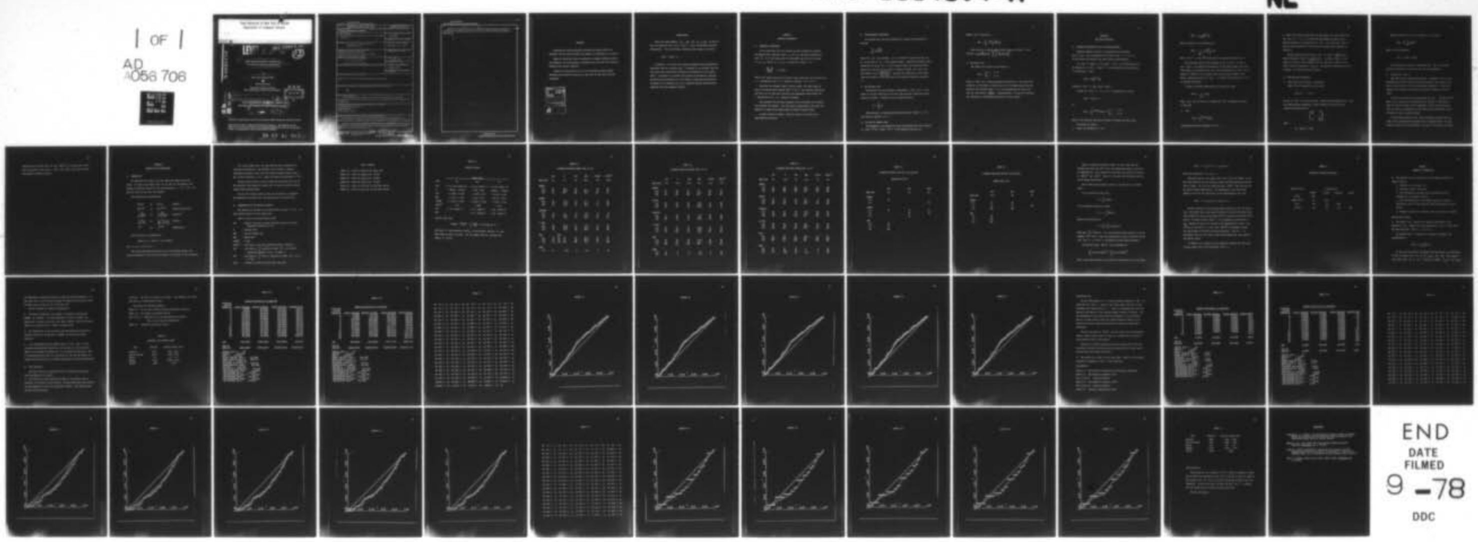
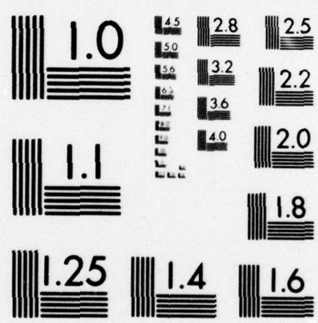
AD-A056 706 STATE UNIV OF NEW YORK AT BUFFALO AMHERST STATISTICA--ETC F/G 12/1
NEW NONPARAMETRIC APPROACH TO THE TWO SAMPLE PROBLEM.(U)
JUL 77 J CARMICHAEL, E PARZEN DAAG29-76-G-0239
UNCLASSIFIED TR-56 ARO-13845.4-M NL

1 OF 1
AD
A056 706



END
DATE
FILMED
9 -78
DDC


1.0
4.5
5.0
5.6
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

State University of New York at Buffalo
Department of Computer Science

AD A056706

LEVEL II

ARO 13845.4-M

18 19

12

6

NEW NONPARAMETRIC APPROACH TO THE TWO SAMPLE PROBLEM*

by

10

Jean-Pierre/Carmichael

[redacted]

Emanuel/Parzen

Statistical Science Division
State University of New York at Buffalo

15

DAAG29-76-G-0239

14

GRANT TECHNICAL REP[redacted]T ~~NO. ARO-2~~ TR-

STATISTICAL SCIENCE DIVISION REPORT NO. 56

9

11 Jul[redacted]77 12 53 p.

*Research supported by the Army Research Office (Grant DA AG29-76-0239).

Approved for public release; distribution unlimited. The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ~~Technical Report No. ARO-2~~	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) New Nonparametric Approach to the Two-Sample Problem		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Jean-Pierre Carmichael and Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) DA AG29-76-G-0239
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistical Science Division State University of New York at Buffalo Amherst, New York 14226		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE July, 1977 13. NUMBER OF PAGES 50
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonparametric Statistics
Rank Statistics
Two-Sample Tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Following the unified approach introduced by Parzen (1977), we one can reexamine the two-sample problem with emphasis on differences of location.

Based on simulations ~~that we performed~~, ~~we compare~~ the different classical parametric and non-parametric procedures are compared with some new test statistics, obtained from Parzen's approach.

continued next page

next page

78 07 31 032

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Finally we illustrate the use of ~~our~~ the two-sample problem package is illustrated including new graphical procedures on some sets of data taken from the literature.

ABSTRACT

Following the unified approach introduced by Parzen (1977), we reexamine the two-sample problem with emphasis on differences of location.

Based on simulations that we performed, we compare different classical parametric and non-parametric procedures with some new test statistics obtained from Parzen's approach.

Finally we illustrate the use of our two-sample problem package including new graphical procedures on some sets of data taken from the literature.

INTRODUCTION

Given two random samples $(X_1,\ldots,X_m)$ and $(Y_1,\ldots,Y_n)$, we want to test the hypothesis that $F_X(\cdot) = F_Y(\cdot)$. There are different possible alternatives. Here we are mostly concerned about change of location:

$$F_Y(x) = F_X(x - \mu) \quad .$$

In Chapter 1, we review the classical parametric and non-parametric procedures that are currently used. In Chapter 2, we introduce some new test statistics obtained from Parzen's new formulation of the problem (1977). In Chapter 3, we present the results of simulations comparing these different procedures on a wide range of underlying distributions. In Chapter 4, we document the use of a computer package developed here, including some new graphical displays.

CHAPTER 1

CLASSICAL PROCEDURES

A. PARAMETRIC PROCEDURES

If we assume that both the X-sample and the Y-sample are normally distributed with respective means μ_1 and μ_2 and common unknown variance σ^2, it is well known that the likelihood ratio test for testing $H : \mu_1 = \mu_2$ versus $K : \mu_1 \neq \mu_2$ is given by: reject H if

$$\frac{|\bar{X} - \bar{Y}|}{s_{\bar{X}-\bar{Y}}} \quad \text{is large.}$$

This is the common two-sided two-sample t-test, where the test statistic has a t distribution with $N-2$ degrees of freedom $(N = n + m)$.

Even when the original samples are not normal, one might argue by way of the Central Limit Theorem that $\bar{X}$ and $\bar{Y}$ are normally distributed and still use the same test statistic with approximate distribution the t distribution with $N-2$ degrees of freedom.

The procedure has also been extended to the case where the variances are different and unknown. One still gets an approximate t-test with the degrees of freedom being approximated by Welch's formula (1949).

It will be seen in Chapter 3 that the t-test is as good as the approximations involved.

B. NON-PARAMETRIC PROCEDURES

We consider only rank tests generated by linear rank statistics of the form

$$\sum_{i=1}^{N} c_i J\left(\frac{R_i}{N}\right)$$

where the c_i's are constants, $J(\cdot)$ is called the score function, and R_i is the rank of X_i in the combined sample. Chernoff and Savage (1958) showed how to choose $J(\cdot)$ in the situation where the underlying distributions are the same up to a shift in location $\left(F_0(x) \text{ and } F_0(x-\theta)\right)$. Then taking $J(u) = \frac{-f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}$ maximizes the efficiency of the linear rank test relative to the test based on the maximum likelihood estimator $\hat{\theta}$.

1. The Wilcoxon Test

The Wilcoxon rank sum statistic corresponds to $J(u) = 2u - 1$; that makes it the most efficient of the rank tests when the underlying distribution is logistic. Actually we use a modified version

$$W = \sum_{i=1}^{m} R_i$$

This statistic is asymptotically normal with mean $\frac{1}{2} m(m + n + 1)$ and variance $\frac{1}{12} mn(m + n + 1)$.

2. The Van der Waerden Test

Corresponding to an underlying normal distribution the score function is $J(u) = \Phi^{-1}(u)$, where $\Phi^{-1}(\cdot)$ is the quantile function of a

Normal (0,1) distribution.

$$VDW = \sum_{i=1}^{m} \Phi^{-1}\left(\frac{R_i}{m+n+1}\right)$$

This statistic is again asymptotically normal with mean 0 and variance $\frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} \left[\Phi^{-1}\left(\frac{i}{m+n+1}\right)\right]^2$.

3. The Median Test

The median test statistic corresponds to

$$J(u) = \begin{cases} -1 , & u < \frac{1}{2} \\ 1 , & u > \frac{1}{2} \end{cases}$$

which is "best" for a double-exponential distribution. The median test statistic is the number of observations in the X-sample lying above the median of the combined sample. It is also asymptotically normal with mean $\frac{m}{2}$ and variance $\frac{mn}{4(m+n)}$ (approximately). We use the correction for continuity in determining approximate critical values.

CHAPTER 2

NEW TEST STATISTICS

A. PARZEN'S FORMULATION OF THE TWO-SAMPLE PROBLEM

Parzen's approach consists in transforming the hypothesis $F_X(\cdot) = F_Y(\cdot)$ into a hypothesis of the form $D(u) = u$, $0 \le u \le 1$. In this report we consider only the following transformation.

Let $H(x) = \lambda F_X(x) + (1-\lambda) F_Y(x)$. $H(\cdot)$ is the distribution of the mixture of X and Y, where λ is the proportion coming from the X distribution. Then,

$$D(u) = F_X\left(H^{-1}(u)\right)$$

satisfies $D(u) = u$, when $F_X(\cdot) = F_Y(\cdot)$.

To test for $D(u) = u$, $0 \le u \le 1$, is equivalent to testing

$$d(u) = D'(u) \equiv 1$$

or

$$\varphi(v) = \int_0^1 e^{2\pi i t v}\, dD(t) = \begin{cases} 1, & v = 0 \\ 0, & v \neq 0 \end{cases}$$

which is the familiar time series problem of testing for white noise.

We proceed as follows:

1. Form a raw estimator of $D(\cdot)$:

$$\tilde{D}(u) = F_{X,m}\left(H_N^{-1}(u)\right)$$

This is computed in the following way

$$\tilde{D}(u) = \sum_{\frac{j}{N} \leq u} w(j) \Big/ \sum_{j=1}^{N} w(j)$$

where $w(j) = 1$ if the $j^{\underline{th}}$ observation in the pooled sample is an X .

This procedure can be easily adapted in case of ties. In such a case, $H_N(\cdot)$, the empirical distribution of the pooled sample, will have some jumps of size greater than $1/N$, but the definition of $\tilde{D}(u)$ is the number of elements in the X-sample that are less than or equal to the $100\,u\,\%$ empirical quantile of $H_N(\cdot)$. Of course, the definition of $w(\cdot)$ is changed accordingly.

In case of censored observations, the definition reads

$$\tilde{D}(u) = \sum_{t_j \leq u} w(t_j)$$

where $\{t_j\}$ are the points of increase and $w(\cdot)$ represents the value of the jump.

2. Form

$$\tilde{\phi}(v) = \int_0^1 e^{2\pi i t v} \, d\tilde{D}(t)$$

the empirical Fourier transform of $\tilde{D}(\cdot)$.

3. There are several possibilities at this stage. We could either test for $\varphi(v) = 0$, $v \neq 0$, or, using the autoregressive method, form a smooth estimator of the derivative of $D(\cdot)$, $d(\cdot)$ and test $d(u) = 1$. Each of these alternatives could be done in several ways, explored in Section B.

Another possible definition for $D(\cdot)$ is $D_1(u) = F_X\left(F_Y^{-1}(u)\right)$ to which corresponds $\tilde{D}_1(u) = F_{X,m}\left(F_{Y,n}^{-1}(u)\right)$. In words, $\tilde{D}_1(u)$ is the number of elements in the X-sample that are less than or equal to the $100\,u\,\%$ empirical quantile of $F_{Y,n}(\cdot)$, the empirical distribution of the Y-sample. The other steps are the same.

B. SOME NEW TEST STATISTICS

1. Tests based on the Fourier coefficients

Under the null hypothesis, we have that

$$\varphi(v) = 0 \ , \quad v \neq 0 \ .$$

We may use $\tilde{\varphi}(\cdot)$ as a test statistic. Under the null hypothesis, the $\tilde{\varphi}(v)$ are asymptotically independent complex Gaussian with mean zero and variance-covariance matrix

$$\begin{pmatrix} \frac{1}{2N} & 0 \\ 0 & \frac{1}{2N} \end{pmatrix} .$$

Then,

$$2N \cdot |\tilde{\varphi}(v)|^2 \sim \chi^2(2)$$

Another possibility is to use statistics of the form

$$A(m) = \frac{1}{m} \sum_{v=1}^{m} |\tilde{\varphi}(v)|^2 .$$

Under the null hypothesis,

$$2N \cdot m \cdot A(m) \sim \chi^2(2m)$$

In our empirical studies, we discovered that $A(1)$ was the most sensitive statistics to deviations from the null hypothesis.

2. Testing for $d(u) \equiv 1$

Here also we have several possibilities. Estimating $d(\cdot)$ by the autoregressive method transforms the null hypothesis into choosing order zero as the best order for the autoregressive estimator. We have used Parzen's CAT criterion and Akaike's criterion to decide on the best order.

We have also looked at some functionals of $d(\cdot)$: $\int_0^1 (\log d(u))^2 du$ which is zero under the null hypothesis and $\int_0^1 (d(u) - 1) \log d(u)\, du$ which is also zero under the null hypothesis. These last functionals were not tested so extensively as the other proposed statistics, and so we report on them in separate tables.

It was found empirically that these functionals increase with the order of the autoregressive estimator used in computing them. For this reason, we have tried two approaches: one was to look only at the value

obtained from the first order (as with $|\tilde{\varphi}(1)|^2$), the other was to subtract a function of the order k (like $(2K + 2)/m$) and find out where the maximum or minimum occurred.

CHAPTER 3

RESULTS OF THE SIMULATIONS

A. METHODOLOGY

We restricted the study to the case where both sample sizes are equal. For each of the sample sizes 10, 20 and 50 we generated 100 samples at different values for the shift parameter Δ : 0.5 , 1.0 , 2.0 . The null case was done with 200 samples.

The distributions involved were:

$N(0,1)$	vs.	$N(\Delta,1)$	(normal)
$\frac{1}{2} e^{-\lvert x \rvert}$	vs.	$\frac{1}{2} e^{-\lvert x-\Delta \rvert}$	(double-exponential)
$\frac{e^{x}}{(1+e^{x})^{2}}$	vs.	$\frac{e^{x-\Delta}}{(1+e^{x-\Delta})^{2}}$	(logistic)
$\frac{1}{\pi(1+x^{2})}$	vs.	$\frac{1}{\pi\left(1+(x-\Delta)^{2}\right)}$	(Cauchy)
e^{-x}	vs.	$e^{-(x-\Delta)}$	(exponential)

We also looked at contamination:

$$N(0,1) \text{ vs. } p \cdot N(0,1) + (1-p)\, N(0,9)$$

for $p = 0.1$, 0.2 and 0.5 .

Note that contamination does not fit in the location problem, but we were determined to test the new procedure in all kinds of test situations.

For a given sample size, the same seed was used to generate the different distributions. The procedure was as follows: generate exponential deviates; obtain from them ordered uniform deviates and, to get a given distribution $F(\cdot)$, apply the inverse transformation $F^{-1}(\cdot)$.

We have used the exact critical values for two-sided tests for all the classical tests except for sample size 50 where we used the corresponding normal approximation.

For the CAT criterion and the new test statistics, we estimated the appropriate quantiles from 200 replications of the null case.

B. PRESENTATION OF THE EMPIRICAL RESULTS

The tables give the power of the given tests at level $\alpha = .05$. We have produced tables for each sample size.

Here is a list of the abbreviations used:

CAT : Parzen's criterion to choose the best order for the autoregressive estimator of $d(\cdot)$

W : Wilcoxon test

VDW : Van der Waerden test

MED : Median test

T-TEST : t-test

$|\tilde{\varphi}(1)|^2$: test based on the first empirical Fourier coefficient

LG2 : test based on $\int_0^1 \log \hat{d}(u)\,du$ where $\hat{d}(\cdot)$ is the autoregressive estimator of $d(\cdot)$ of order 1

DIV : test based on $\int_0^1 (\hat{d}(u) - 1) \log \hat{d}(u)\,du$ where $\hat{d}(\cdot)$ is as in LG2

DIVC : criterion to choose the best order using DIV.

List of Tables

TABLE 3.1

CRITICAL VALUES

Test	Sample Size 10	20	50
CAT	≤ -1.1, for orders 1,2	≤ -1.05, orders $\to 3$	≤ -1.02, orders $\to 9$
W	≤ 78.75, ≥ 131.25	≤ 337, ≥ 483	≤ 2240.7, ≥ 2809.3
VDW	≤ -3.88, ≥ 3.88	≤ -5.75, ≥ 5.75	≤ -9.46, ≥ 9.46
MED	≤ 2.17, ≥ 7.83	≤ 6.3, ≥ 13.7	≤ 19.5, ≥ 30.5
T-TEST	≤ -2.101, ≥ 2.101	≤ -2.021, ≥ 2.021	≤ -1.98, ≥ 1.98
$\lvert\tilde{\varphi}(1)\rvert^2$	$\ge .13$	$\ge .056$	$\ge .03$
LG2	$\ge .31$ (order 1)	$\ge .14$ (order 1)	$\ge .05$ (order 1)
DIV		$\ge .115$ (order 1)	$\ge .06$ (order 1)

For the DIVC test:

$$\mathrm{DIV}(k) = \frac{2k+2}{n} - \int_0^1 (\hat{d}_{(k)}(u) - 1) \log \hat{d}_{(k)}(u)\, du \quad ,$$

then find k^* that minimizes $\mathrm{DIV}(k)$; if the minimum $\mathrm{DIV}(k^*) > 0$, the best order is taken to be zero. For the sample size 10 , consider only orders $k = 1,2,3,4$

TABLE 3.2

% CORRECT DECISIONS SAMPLE SIZE 10 - 10

	CAT		W	VDW	MED	T-TEST	$\lvert\tilde{\varphi}(1)\rvert^2$
NULL CASE	94		97	96	96.6	96	93.5
NORMAL							
0.5	11		14.7	14	7	15	10
1.0	27		54.5	54	32.5	57	25
2.0	84		100	100	93.9	100	91
CAUCHY							
0.5	10		3.7	4	4	3	14
1.0	23		19.7	15	15.1	9	29
2.0	55		44.2	34	52.1	21	69
LOGISTIC							
0.5	8		4	5	3.5	4	9
1.0	16		18.7	19	11.4	21	10
2.0	44		69.5	67	51.3	69	42
D. EXP.							
0.5	10		10.2	11	8.7	8	18
1.0	29		38.2	36	31.4	37	30
2.0	73		95.2	89	84.9	87	84
EXP		LG2					
0.5	29	33	36	41	23.6	31	33
1.0	65	71	77	82	61.6	70	71
2.0	96	98	100	100	97.5	97	100
CONT							
0.2	11		3.75	5	3.7	3	14

TABLE 3.3

% CORRECT DECISIONS SAMPLE SIZE 20 - 20

	CAT	W	VDW	MED	T-TEST	$\lvert\tilde{\varphi}(1)\rvert^2$
NULL CASE	95.6	95.6	94.3	95.4	93.6	94.3
NORMAL						
0.5	17	31	31	21	28	24
1.0	45	88	89	67.9	91	61
2.0	99	100	100	100	100	100
CAUCHY						
0.5	11	10	12	12.5	5	24
1.0	37	36	27	44.9	10	65
2.0	88	83	70	88.3	25	95
LOGISTIC						
0.5	5	20	15	7.9	16	15
1.0	26	42	42	30.4	40	32
2.0	68	94	94	84.1	94	86
D. EXP.						
0.5	19	23	21	22.2	17	30
1.0	52	73	67	68.9	59	77
2.0	99	100	99	98.6	100	100
EXP.						
0.5	55	66	70	36.9	40	65
1.0	96	99	98	91.2	92	97
2.0	100	100	100	100	99	100
CONT						
0.1	13	5	6	3.2	8	16
0.2	26	6	7	3.2	7	26

TABLE 3.4

% CORRECT DECISIONS SAMPLE SIZE 50 - 50

	CAT	W	VDW	MED	T-TEST	$\lvert\tilde{\varphi}(1)\rvert^2$
NULL CASE	93	93.5	93	94.5	93	92
NORMAL						
0.5	41	71	73	63	73	38
1.0	92	100	100	98.5	100	89
2.0	100	100	100	100	100	100
CAUCHY						
0.5	33	38	32	45	5	37
1.0	83	75	65	82.5	13	87
2.0	100	100	99	100	25	100
LOGISTIC						
0.5	19	41	38	32.5	38	22
1.0	58	80	79	74.5	78	61
2.0	98	100	100	100	100	98
D. EXP.						
0.5	55	65	59	68	52	59
1.0	95	99	99	98.5	91	97
2.0	100	100	100	100	100	100
EXP.						
0.5	91	96	97	69.5	76	81
1.0	100	100	100	100	100	100
2.0	100	100	100	100	100	100
CONT						
0.1	73	12	20	3	6	20
0.2	95	22	31	4	7	43
0.5	93	24.6	32	4.5	5	62

TABLE 3.5

% CORRECT DECISIONS FOR LG2, DIV AND DIVC

SAMPLE SIZE 20-20

	LG2	DIV	DIVC
NULL CASE	95	95	96
NORMAL			
0.5	19		
1.0			53
CAUCHY			
0.5	18	24	24
1.0		65	52
2.0		95	
EXP			
0.5	58		
1.0	96		
2.0	100		

TABLE 3.6

% CORRECT DECISIONS FOR LG2, DIV AND DIVC

SAMPLE SIZE 50-50

	LG2	DIV	DIVC
NULL CASE	93.5	93	96
NORMAL			
0.5	47	39	
1.0	92	91	88
2.0	100	100	
EXP			
0.5	86		
1.0	100		
1.0	100		

What do we gather from these tables? We note first that the Wilcoxon test does very well in all the comparisons except in the case of contamination. Also, amongst the new tests, none seem to do as well as $|\tilde{\varphi}(1)|^2$ and $|\tilde{\varphi}(1)|^2$ seems to be the best for very heavy tails as in the Cauchy distribution.

Can we find some theoretical reason to justify the use of these tests?

In our notation, we can write

$$W = \int_0^1 u\, d\tilde{D}(u)$$

If we integrate by parts, we obtain

$$W = 1 - \int_0^1 \tilde{D}(u)\, du$$

which can be rewritten as

$$W = \frac{1}{2} + \int_0^1 (u - \tilde{D}(u))\, du \quad .$$

Note that $\int_0^1 (u - \tilde{D}(u))\, du$ is a very intuitive test statistic to use to compare $D(u)$ with u when the alternatives we have in mind are of the form $D(u) < u$ or $D(u) > u$ as they are in the location problem.

On the other hand, $|\tilde{\varphi}(1)|^2$ can be expressed as

$$\left|\int_0^1 \cos 2\pi\, u\, d\tilde{D}(u)\right|^2 + \left|\int_0^1 \sin 2\pi\, u\, d\tilde{D}(u)\right|^2$$

This is the right statistic to use when the alternatives are of the form

$$d(u) = 1 + \theta_1 \cos 2\pi u + \theta_2 \sin 2\pi u$$

and we are testing for $\theta_1 = \theta_2 = 0$.

Parzen has shown in his report (1977) that $\int_0^1 \sin 2\pi u \, d\tilde{D}(u)$ is the best test statistic for the location problem when the underlying distribution is Cauchy. So it is not surprising that $|\tilde{\varphi}(1)|^2$ does very well in the case of Cauchy distribution. Its performance in the other cases depends on how well the alternative can be approximated by the form

$$d(u) = 1 + \theta_1 \cos 2\pi u + \theta_2 \sin 2\pi u \quad .$$

We have examined so far the small sample behavior of our test statistics. What about their large sample behavior and in particular what about their asymptotic relative efficiency? Table 3.7 contains the relevant information. We cannot include $|\tilde{\varphi}(1)|^2$ as its asymptotic distribution is $\chi^2_{(2)}$ whereas the other test statistics are asymptotically normal. But we can see from Tables 2, 3 and 4 that $|\tilde{\varphi}(1)|^2$'s performance follows the same pattern as the ARE of the score function $-\sin 2\pi u$: in decreasing order best for Cauchy, then double-exponential, then logistic and finally normal.

In Chapter 4, we present some new graphical displays that allow one to make visual tests of the hypothesis $D(u) = u$.

TABLE 3.7

ASYMPTOTIC RELATIVE EFFICIENCY

Score Function	Distribution			
	Normal	D.-Exp	Logistic	Cauchy
$\Phi^{-1}(u)$	1.00			
sign (2u - 1)	.64	1.00		
2u - 1	.95	.75	1.00	
$-\sin 2\pi u$	.43	.81	.61	1.00

CHAPTER 4

COMPUTER IMPLEMENTATION

A. The algorithm that we propose for the two-sample problem has a simple structure:

1. Compute a set of weights $W(\cdot)$.

2. Take their Fourier transform.

3. Fit autoregressive schemes using the previous Fourier transform as a covariance function.

4. Plot the partial sum of the weights against the uniform distribution (the 45° line) and the smooth autoregressive distribution.

5. Compute any other test statistic that one may wish to compute.

Step-by-Step Analysis

1. The weights $W(\cdot)$ represent the jumps in the function $\tilde{D}(\cdot)$ estimating $D(\cdot)$; remember the only restriction on $D(\cdot)$ is that under the null hypothesis, $D(u) = u$, $0 \le u \le 1$.

We presently have 3 subroutines to compute the weights, all corresponding to

$$\tilde{D}(u) = F_{X,m}(H_N^{-1}(u)) \quad .$$

In WC , the ties between the samples have been broken by randomization. In WCF , we ignore ties, i.e., we let $F_{X,m}(\cdot)$ and $H_N(\cdot)$ have jumps of size larger than $1/m$ or $1/N$. Finally, in KMPARZ , $F_{X,m}(\cdot)$ and $H_N(\cdot)$

are Kaplan-Meier estimators designed to handle censored observations. In this last case, we also record the number of jumps and the points at which the jumps occur as they are not of the form j/N .

See the listings for complete documentation.

2. The Fourier transform of the weights is computed in subroutines FORIER1 and FORIER2 . The only difference is that in FORIER1 the jumps occur at points of the form j/N while FORIER2 allows for general jumps as in conjunction with KMPARZ (censored data).

3. The coefficients of the different order autoregressive schemes are computed iteratively in subroutine AUTOREG by solving Yule-Walker equations.

4. It is recommended to make several plots of $\tilde{D}(u)$, $D(u) = u$ and the smooth autoregressive estimators, one for each of a few successive orders of the estimator (perhaps up to 5) and study how they vary. Ease of interpretation will come with experience, but one can see whether the shapes change and whether the estimator intersects the uniform distribution.

B. Case Histories.

The first two sets of data that we use to illustrate our package come from Maguire et al (1952).

1. The first set of data represents lengths of time between serious accidents in the mines of Great Britain. We previously found some evidence of non-homogeneity in this data (Carmichael (1976)). This time we split the data into two groups:

1st Group: the first 60 values, 2nd Group: the remaining 49 values (the data is in chronological order).

We present the following exhibits:

Table 4.1: The two sets of data and some descriptive statistics.

Table 4.2: The weights as computed from WC .

Fig. 4.1-4.5: Superposition of the autoregressive estimator, $\tilde{D}(\cdot)$ and the uniform distribution.

Table 4.3: Classical nonparametric tests.

TABLE 4.3

CLASSICAL NONPARAMETRIC TESTS

Test	Observed	Critical Values (95%)
Wilcoxon	2910	2978 - 3621
Van der Waerden	-11.75	-9.85 - 9.85
Median	21	21 - 35
Savage	47.12	50.03 - 69.97
$\lvert\tilde{\varphi}(1)\rvert^2$	.0246	.0275

TABLE 4.1 A

ORDER_STATISTICS_IN_QUARTERS

SEQUENCE WITHIN QUARTILE	FIRST_QUARTER	SECOND_QUARTER	THIRD_QUARTER	FOURTH_QUARI
1	1.0000	50.0000	108.0000	275.0(
2	4.0000	54.0000	113.0000	275.0(
3	11.0000	55.0000	114.0000	286.0(
4	13.0000	58.0000	120.0000	312.0(
5	15.0000	59.0000	123.0000	315.0(
6	15.0000	59.0000	124.0000	326.0(
7	17.0000	61.0000	137.0000	345.0(
8	20.0000	61.0000	151.0000	354.0(
9	22.0000	72.0000	176.0000	361.0(
10	23.0000	78.0000	188.0000	378.0(
11	28.0000	78.0000	189.0000	457.0(
12	31.0000	81.0000	203.0000	467.0(
13	32.0000	93.0000	215.0000	644.0(
14	36.0000	96.0000	217.0000	871.0(
15	48.0000	99.0000	233.0000	1205.0(
SUM	316.0000	1054.0000	2411.0000	6871.0(
SUM OF SQUARES	8828.0000	77788.0000	414757.0000	4105777.0(

SAMPLE SIZE = 60
MEDIAN = 103.500
MEAN = 177.533
INNER FOURTHS MEAN = 115.500
OUTER FOURTHS MEAN = 239.567
VARIANCE = 46034.999
STANDARD DEVIATION = 214.558
INTERQUARTILE RANGE = 227.000
TRIMEAN = 132.500
GASTWIRTH'S ESTIMATE = 115.500
.05-WINSORIZED MEAN = 147.933
.05-TRIMMED MEAN = 146.593
.10-WINSORIZED MEAN = 141.017
.10-TRIMMED MEAN = 136.896
.25-WINSORIZED MEAN = 123.783
.25-TRIMMED MEAN = 115.500

TABLE 4.1 B

ORDER_STATISTICS_IN_QUARTERS

SEQUENCE WITHIN QUARTILE	FIRST_QUARTER	SECOND_QUARTER	THIRD_QUARTER	FOURTH_QUART
1	4.0000	72.0000	217.0000	338.00
2	7.0000	75.0000	224.0000	348.00
3	18.0000	120.0000	228.0000	364.00
4	19.0000	129.0000	255.0000	369.00
5	19.0000	131.0000	271.0000	390.00
6	20.0000	145.0000	275.0000	498.00
7	29.0000	156.0000	291.0000	517.00
8	37.0000	171.0000	312.0000	566.00
9	47.0000	182.0000	326.0000	745.00
10	49.0000	195.0000	329.0000	1312.00
11	54.0000	208.0000	330.0000	1357.00
12	66.0000	217.0000	336.0000	1613.00
13				1630.00
SUM	369.0000	1801.0000	3394.0000	10047.00
SUM OF SQUARES	15603.0000	295115.0000	981678.0000	10368241.00

SAMPLE SIZE = 49
MEDIAN = 217.000
MEAN = 318.592
INNER FOURTHS MEAN = 216.458
OUTER FOURTHS MEAN = 416.640
VARIANCE = 149731.247
STANDARD DEVIATION = 386.951
INTERQUARTILE RANGE = 272.000
TRIMEAN = 209.500
GASTWIRTH'S ESTIMATE = 223.900
.05-WINSORIZED MEAN = 280.245
.05-TRIMMED MEAN = 274.600
.10-WINSORIZED MEAN = 243.735
.10-TRIMMED MEAN = 235.393
.25-WINSORIZED MEAN = 204.959
.25-TRIMMED MEAN = 221.320

TABLE 4.2

W(1) = 1.	W(2) = 0.	W(3) = 1.	W(4) = 0.	W(5) = 1.
W(6) = 1.	W(7) = 1.	W(8) = 1.	W(9) = 1.	W(10) = 0.
W(11) = 0.	W(12) = 0.	W(13) = 1.	W(14) = 0.	W(15) = 1.
W(16) = 1.	W(17) = 1.	W(18) = 0.	W(19) = 1.	W(20) = 1.
W(21) = 1.	W(22) = 0.	W(23) = 0.	W(24) = 1.	W(25) = 0.
W(26) = 1.	W(27) = 0.	W(28) = 1.	W(29) = 1.	W(30) = 1.
W(31) = 1.	W(32) = 1.	W(33) = 1.	W(34) = 1.	W(35) = 0.
W(36) = 0.	W(37) = 1.	W(38) = 0.	W(39) = 1.	W(40) = 1.
W(41) = 1.	W(42) = 1.	W(43) = 1.	W(44) = 1.	W(45) = 1.
W(46) = 1.	W(47) = 1.	W(48) = 1.	W(49) = 0.	W(50) = 1.
W(51) = 1.	W(52) = 0.	W(53) = 0.	W(54) = 1.	W(55) = 0.
W(56) = 1.	W(57) = 0.	W(58) = 0.	W(59) = 1.	W(60) = 0.
W(61) = 1.	W(62) = 1.	W(63) = 0.	W(64) = 1.	W(65) = 0.
W(66) = 1.	W(67) = 0.	W(68) = 1.	W(69) = 0.	W(70) = 0.
W(71) = 0.	W(72) = 1.	W(73) = 0.	W(74) = 0.	W(75) = 1.
W(76) = 0.	W(77) = 1.	W(78) = 1.	W(79) = 0.	W(80) = 1.
W(81) = 0.	W(82) = 1.	W(83) = 1.	W(84) = 0.	W(85) = 0.
W(86) = 0.	W(87) = 0.	W(88) = 0.	W(89) = 1.	W(90) = 0.
W(91) = 1.	W(92) = 1.	W(93) = 0.	W(94) = 0.	W(95) = 1.
W(96) = 0.	W(97) = 1.	W(98) = 1.	W(99) = 0.	W(100) = 0.
W(101) = 0.	W(102) = 1.	W(103) = 0.	W(104) = 1.	W(105) = 1.
W(106) = 0.	W(107) = 0.	W(108) = 0.	W(109) = 0.	W(

FIGURE 4.1

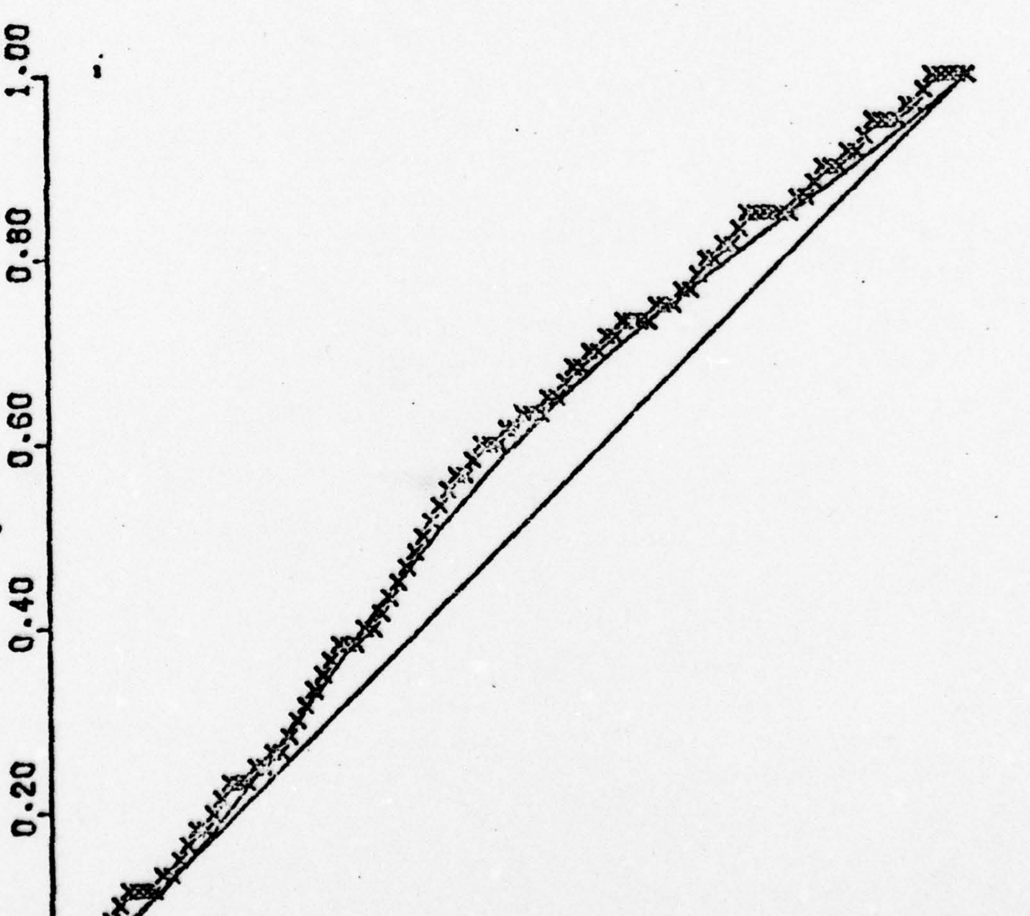


FIGURE 4.2

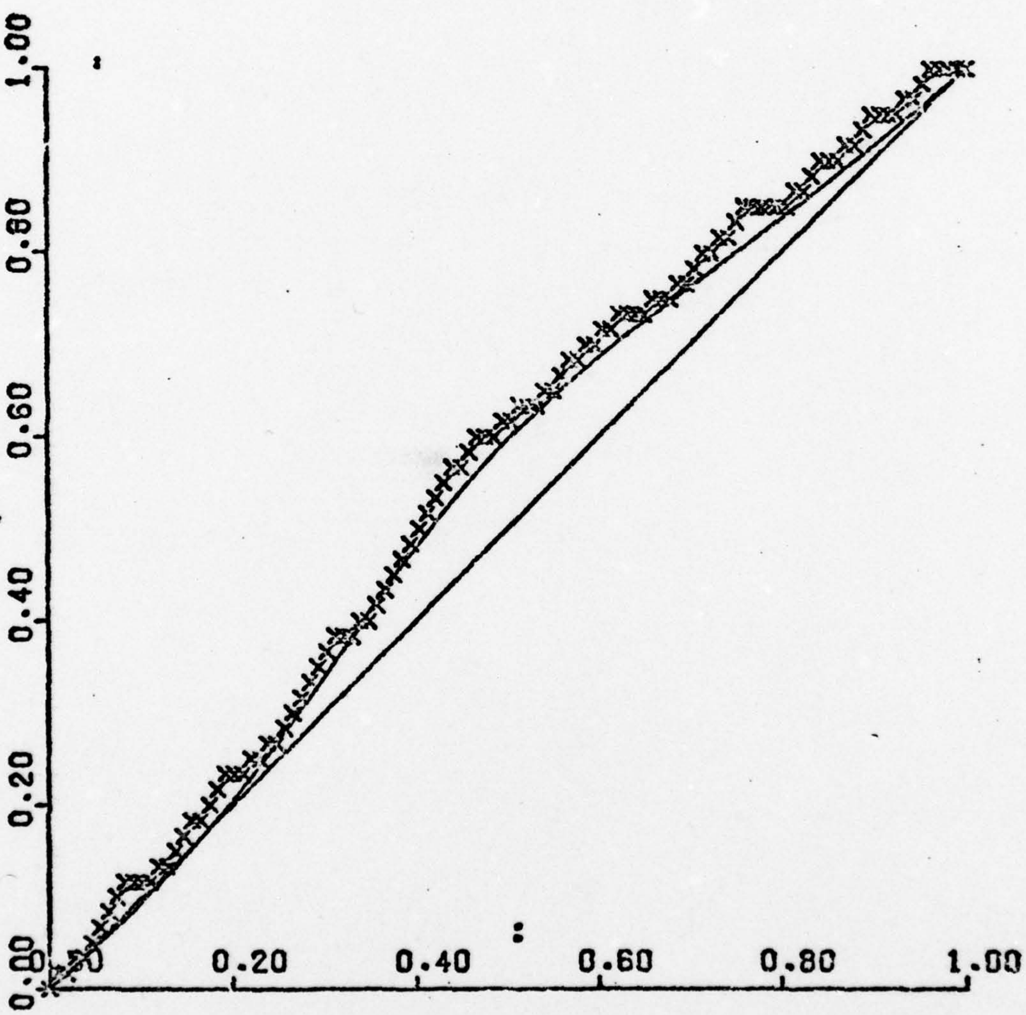


FIGURE 4.3

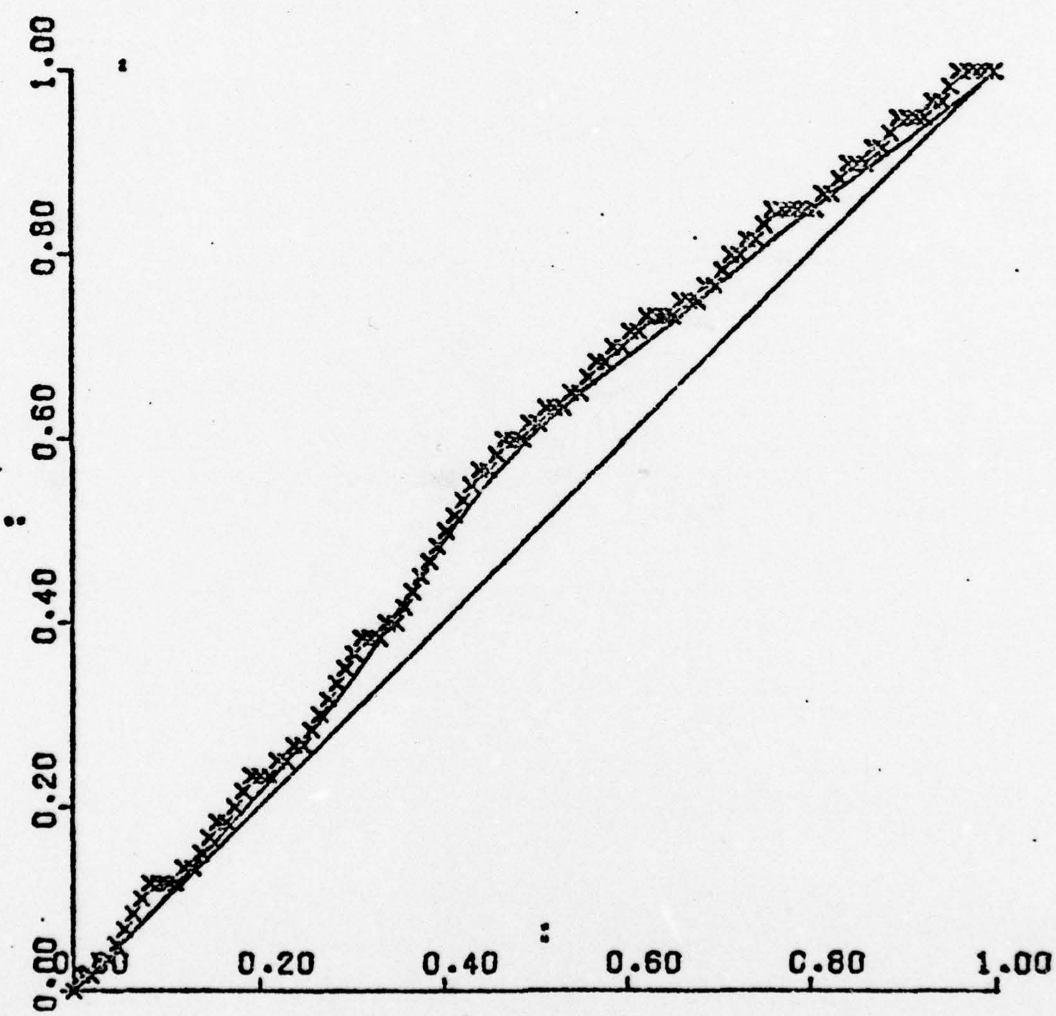


FIGURE 4.4

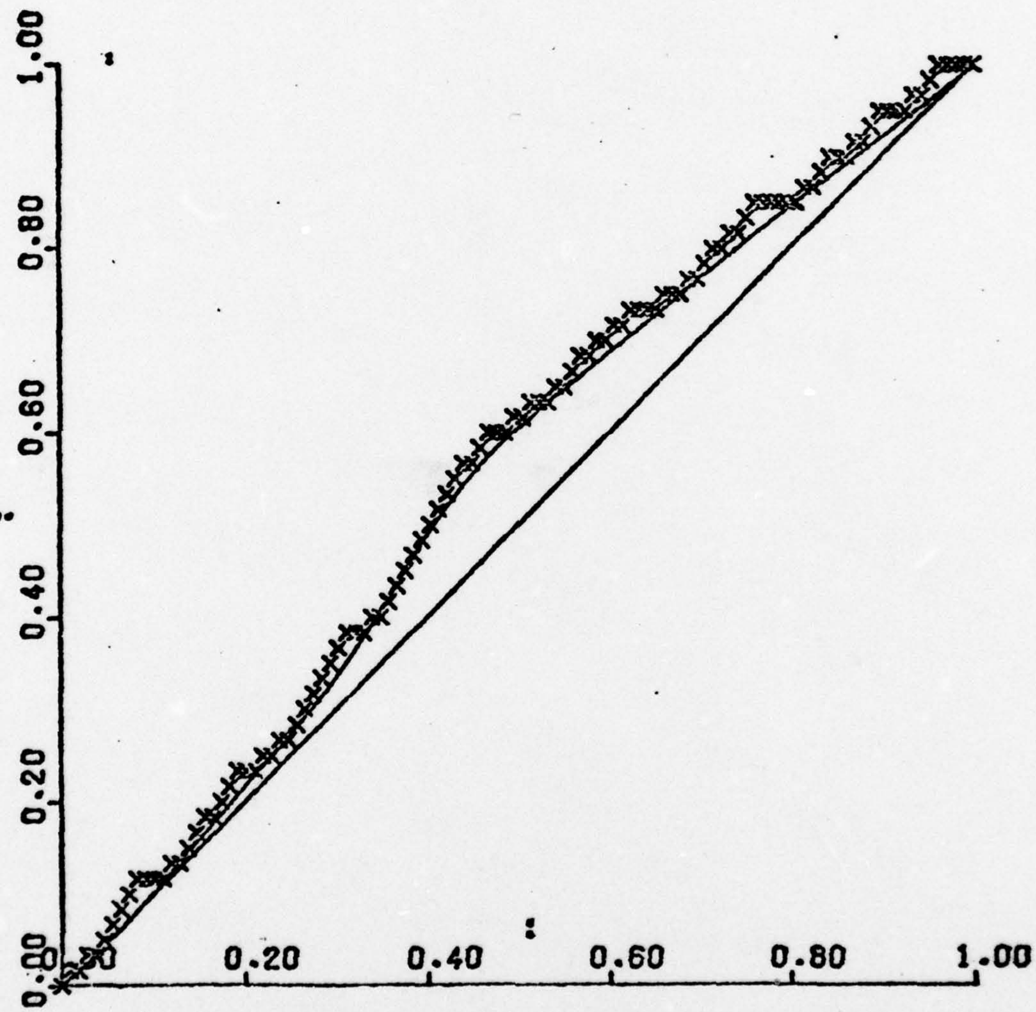
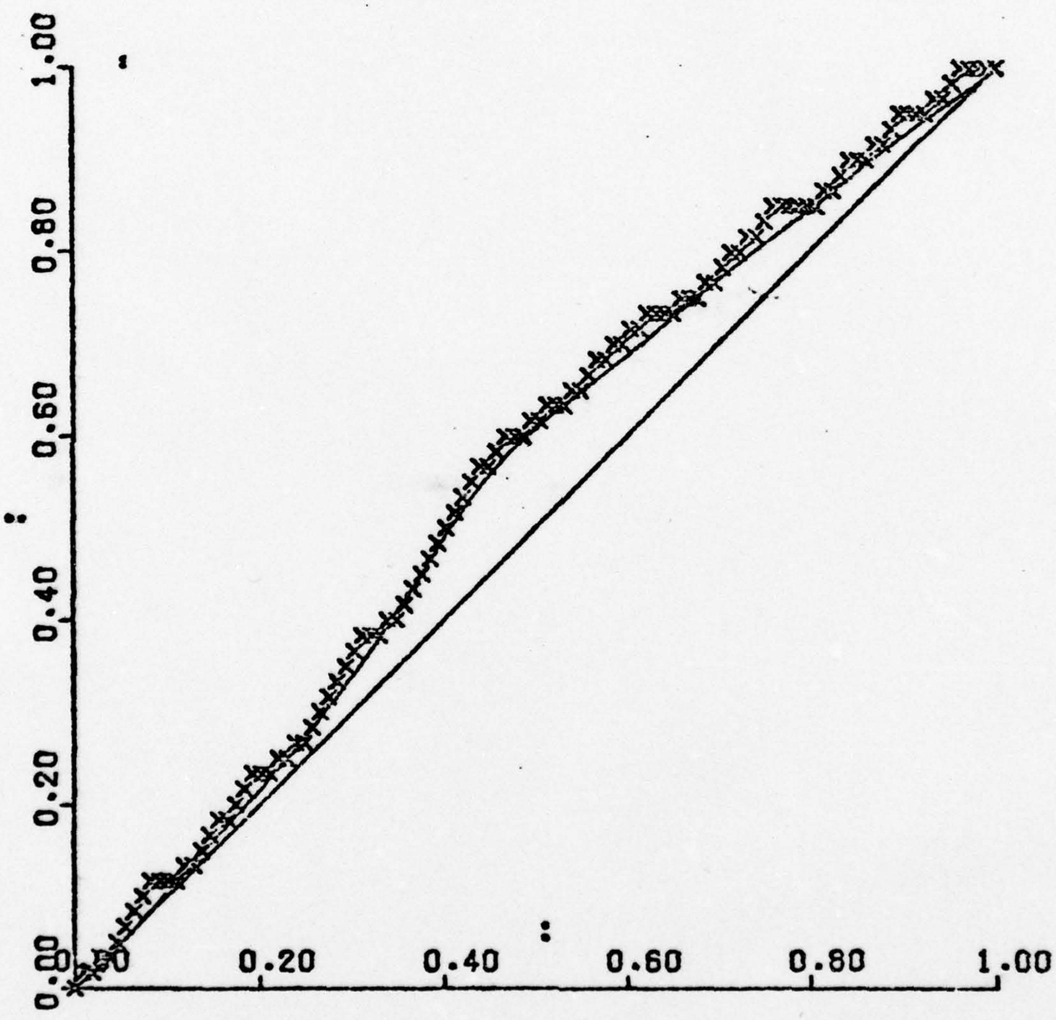


FIGURE 4.5



Interpretations:

We note from Figures 4.1 - 4.5 that the smooth estimator of $\tilde{D}(\cdot)$ is above the line $D(u) = u$ almost in the entire range, and this is more pronounced with large values of u. This is in agreement with our estimation of the density of the combined samples referred to earlier. All the nonparametric tests would reject the hypothesis of no difference. This is in sharp contrast with the results obtained by Maguire et al, where all the tests they performed failed to detect any significant difference.

The new test based on $|\tilde{\varphi}(1)|^2$ does not reject the null hypothesis either. What we need to give an answer is a combination of looking at test statistics and at the pictures.

From this, we would conclude that the two groups don't differ for low values, but that the second group has experienced the longer interarrival times with higher probability.

2. The second set of data is of the same type: length of time between accidents in divisions 3 and 7 of the same mine.

The exhibits:

Table 4.4: The two sets of data and some descriptive statistics.

Table 4.5: The weights as computed in WC .

Fig. 4.6-4.10: Pictures obtained.

Table 4.6: The weights as computed in WCF .

Fig. 4.11-4.15: Pictures obtained.

Table 4.7: Classical nonparametric tests.

TABLE 4.4 A

ORDER STATISTICS IN QUARTERS

SEQUENCE WITHIN QUARTILE	FIRST QUARTER	SECOND QUARTER	THIRD QUARTER	FOURTH QUAR
1	0.0000	3.0000	6.0000	11.(
2	0.0000	3.0000	6.0000	11.(
3	0.0000	4.0000	6.0000	11.0
4	1.0000	4.0000	7.0000	12.0
5	1.0000	5.0000	7.0000	13.0
6	1.0000	5.0000	8.0000	13.0
7	1.0000	5.0000	8.0000	13.0
8	1.0000	5.0000	8.0000	13.0
9	2.0000	5.0000	9.0000	17.0
10	2.0000			18.0
SUM	9.0000	39.0000	65.0000	132.0
SUM OF SQUARES	13.0000	175.0000	479.0000	1796.0

SAMPLE SIZE = 38
MEDIAN = 5.500
MEAN = 6.447
INNER FOURTHS MEAN = 5.778
OUTER FOURTHS MEAN = 7.050
VARIANCE = 23.876
STANDARD DEVIATION = 4.886
INTERQUARTILE RANGE = 9.000
TRIMEAN = 6.000
GASTWIRTH'S ESTIMATE = 5.800
.05-WINSORIZED MEAN = 5.974
.05-TRIMMED MEAN = 6.306
.10-WINSORIZED MEAN = 5.921
.10-TRIMMED MEAN = 6.156
.25-WINSORIZED MEAN = 5.816
.25-TRIMMED MEAN = 5.850

TABLE 4.4 B

ORDER_STATISTICS_IN_QUARTERS

SEQUENCE WITHIN QUARTILE	FIRST_QUARTER	SECOND_QUARTER	THIRD_QUARTER	FOURTH_QUA
1	0.0000	1.0000	2.0000	6.
2	0.0000	1.0000	2.0000	6.
3	0.0000	1.0000	3.0000	6.
4	0.0000	1.0000	3.0000	7.
5	0.0000	1.0000	3.0000	7.
6	0.0000	1.0000	4.0000	8.
7	0.0000	2.0000	4.0000	8.
8	0.0000	2.0000	4.0000	9.
9	0.0000	2.0000	5.0000	10.
10	0.0000	2.0000	5.0000	13.
11	0.0000	2.0000	5.0000	14.
12	1.0000	2.0000	5.0000	14.
13	1.0000	2.0000	5.0000	15.
14	1.0000	2.0000	6.0000	17.
15				24.
SUM	3.0000	22.0000	56.0000	164.
SUM OF SQUARES	3.0000	38.0000	244.0000	2166.

SAMPLE SIZE = 57
MEDIAN = 2.000
MEAN = 4.298
INNER FOURTHS MEAN = 2.786
OUTER FOURTHS MEAN = 5.759
VARIANCE = 24.963
STANDARD DEVIATION = 4.996
INTERQUARTILE RANGE = 5.000
TRIMEAN = 2.750
GASTWIRTH'S ESTIMATE = 2.600
.05-WINSORIZED MEAN = 3.842
.05-TRIMMED MEAN = 3.849
.10-WINSORIZED MEAN = 3.737
.10-TRIMMED MEAN = 3.426
.25-WINSORIZED MEAN = 3.070
.25-TRIMMED MEAN = 2.897

TABLE 4.5

W(1) = 0.	W(2) = 0.	W(3) = 0.	W(4) = 0.	W(5) = 0.
W(6) = 0.	W(7) = 0.	W(8) = 1.	W(9) = 0.	W(10) = 0.
W(11) = 0.	W(12) = 0.	W(13) = 1.	W(14) = 1.	W(15) = 0.
W(16) = 0.	W(17) = 0.	W(18) = 0.	W(19) = 0.	W(20) = 0.
W(21) = 1.	W(22) = 0.	W(23) = 0.	W(24) = 0.	W(25) = 1.
W(26) = 1.	W(27) = 1.	W(28) = 1.	W(29) = 0.	W(30) = 0.
W(31) = 0.	W(32) = 0.	W(33) = 1.	W(34) = 0.	W(35) = 0.
W(36) = 0.	W(37) = 0.	W(38) = 0.	W(39) = 0.	W(40) = 1.
W(41) = 0.	W(42) = 0.	W(43) = 1.	W(44) = 0.	W(45) = 1.
W(46) = 0.	W(47) = 0.	W(48) = 0.	W(49) = 1.	W(50) = 1.
W(51) = 0.	W(52) = 0.	W(53) = 1.	W(54) = 0.	W(55) = 0.
W(56) = 0.	W(57) = 1.	W(58) = 1.	W(59) = 1.	W(60) = 1.
W(61) = 0.	W(62) = 0.	W(63) = 1.	W(64) = 0.	W(65) = 0.
W(66) = 1.	W(67) = 1.	W(68) = 0.	W(69) = 1.	W(70) = 0.
W(71) = 1.	W(72) = 1.	W(73) = 0.	W(74) = 0.	W(75) = 1.
W(76) = 1.	W(77) = 1.	W(78) = 0.	W(79) = 0.	W(80) = 1.
W(81) = 1.	W(82) = 1.	W(83) = 1.	W(84) = 1.	W(85) = 0.
W(86) = 1.	W(87) = 1.	W(88) = 1.	W(89) = 0.	W(90) = 0.
W(91) = 0.	W(92) = 0.	W(93) = 1.	W(94) = 1.	W(95) = 0.

FIGURE 4.6

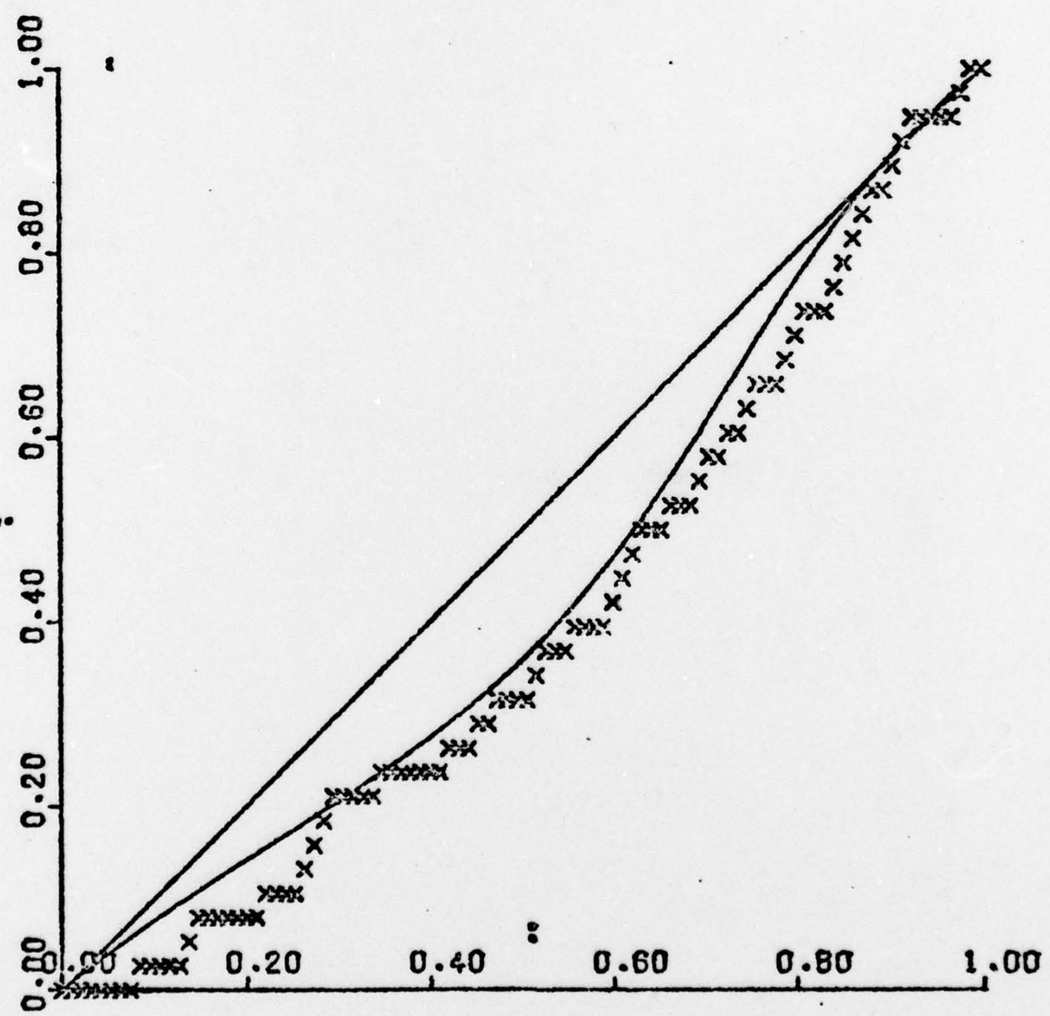


FIGURE 4.7

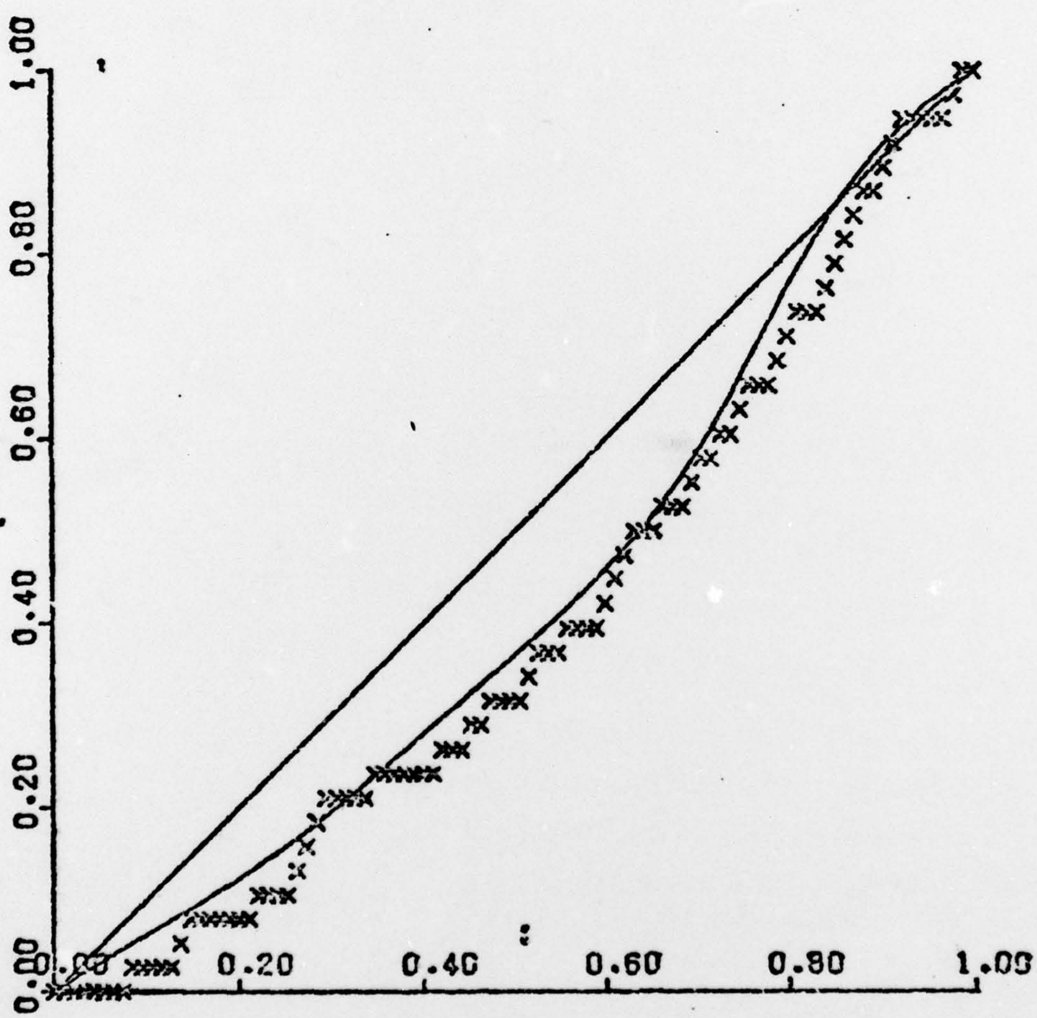


FIGURE 4.8

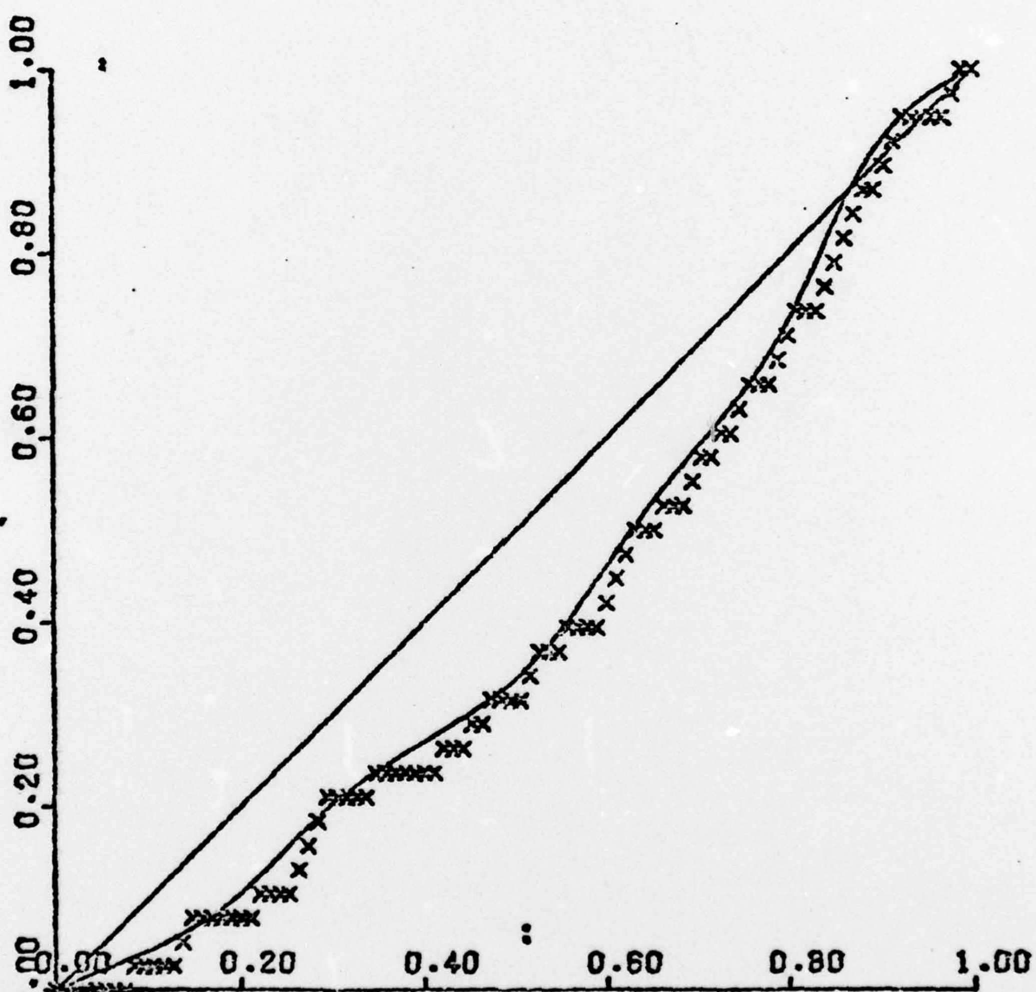


FIGURE 4.9

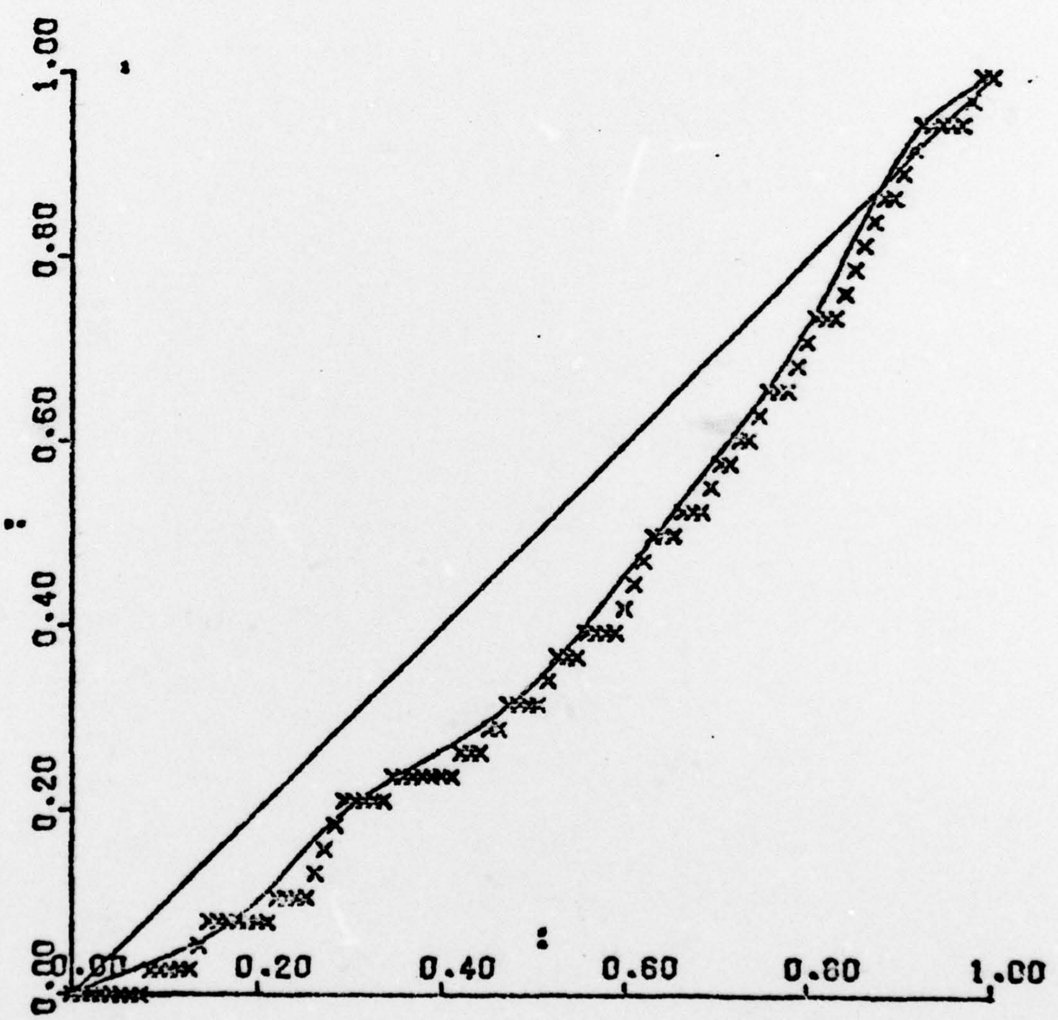


FIGURE 4.10

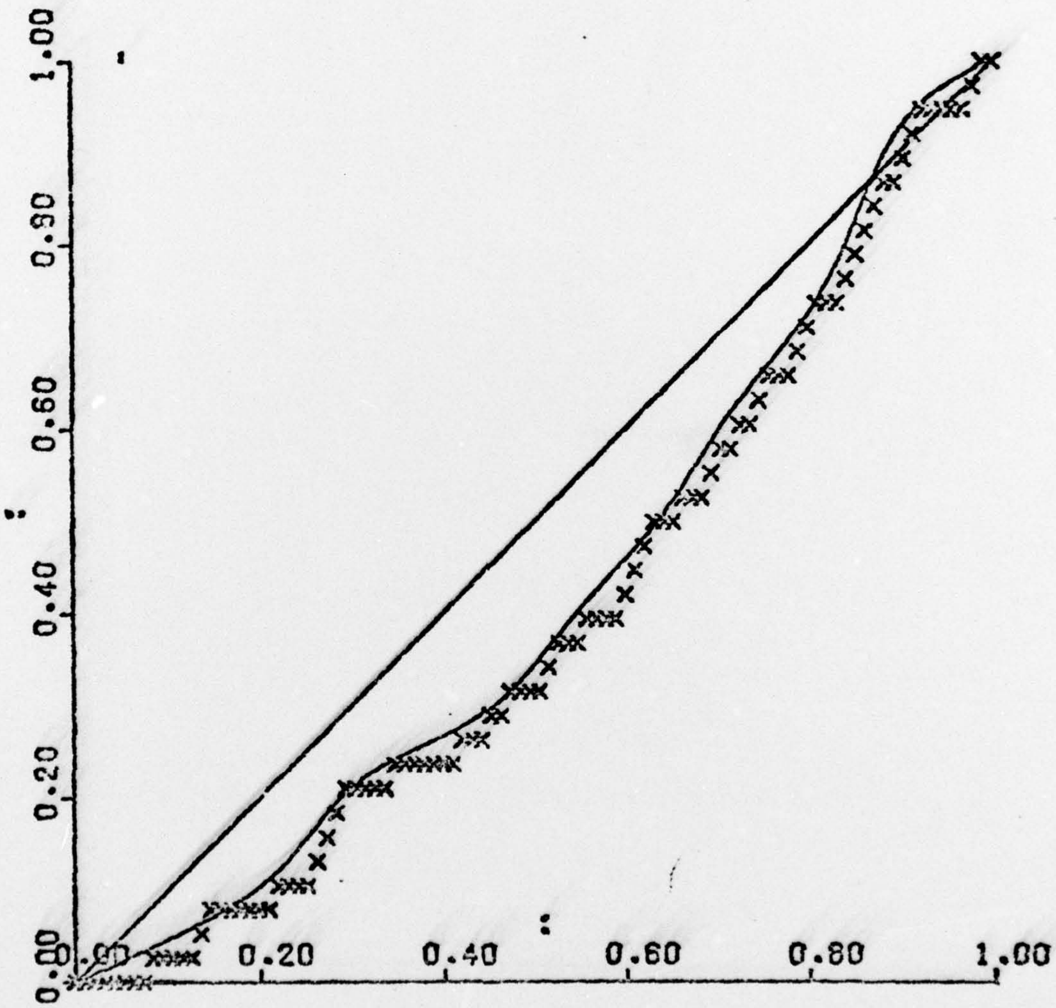


TABLE 4.6

W(1) = 3.	W(2) = 0.	W(3) = 0.	W(4) = 0.	W(5) = 0.
W(6) = 0.	W(7) = 0.	W(8) = 0.	W(9) = 0.	W(10) = 0.
W(11) = 0.	W(12) = 0.	W(13) = 0.	W(14) = 0.	W(15) = 5.
W(16) = 0.	W(17) = 0.	W(18) = 0.	W(19) = 0.	W(20) = 0.
W(21) = 0.	W(22) = 0.	W(23) = 0.	W(24) = 0.	W(25) = 0.
W(26) = 0.	W(27) = 0.	W(28) = 0.	W(29) = 2.	W(30) = 0.
W(31) = 0.	W(32) = 0.	W(33) = 0.	W(34) = 0.	W(35) = 0.
W(36) = 0.	W(37) = 0.	W(38) = 0.	W(39) = 0.	W(40) = 0.
W(41) = 2.	W(42) = 0.	W(43) = 0.	W(44) = 0.	W(45) = 0.
W(46) = 2.	W(47) = 0.	W(48) = 0.	W(49) = 0.	W(50) = 0.
W(51) = 5.	W(52) = 0.	W(53) = 0.	W(54) = 0.	W(55) = 0.
W(56) = 0.	W(57) = 0.	W(58) = 0.	W(59) = 0.	W(60) = 0.
W(61) = 3.	W(62) = 0.	W(63) = 0.	W(64) = 0.	W(65) = 0.
W(66) = 0.	W(67) = 0.	W(68) = 2.	W(69) = 0.	W(70) = 0.
W(71) = 0.	W(72) = 2.	W(73) = 0.	W(74) = 0.	W(75) = 0.
W(76) = 0.	W(77) = 1.	W(78) = 0.	W(79) = 0.	W(80) = 3.
W(81) = 0.	W(82) = 0.	W(83) = 1.	W(84) = 4.	W(85) = 0.
W(86) = 0.	W(87) = 0.	W(88) = 0.	W(89) = 0.	W(90) = 0.
W(91) = 0.	W(92) = 1.	W(93) = 0.	W(94) = 1.	W(95) = 0.

FIGURE 4.11

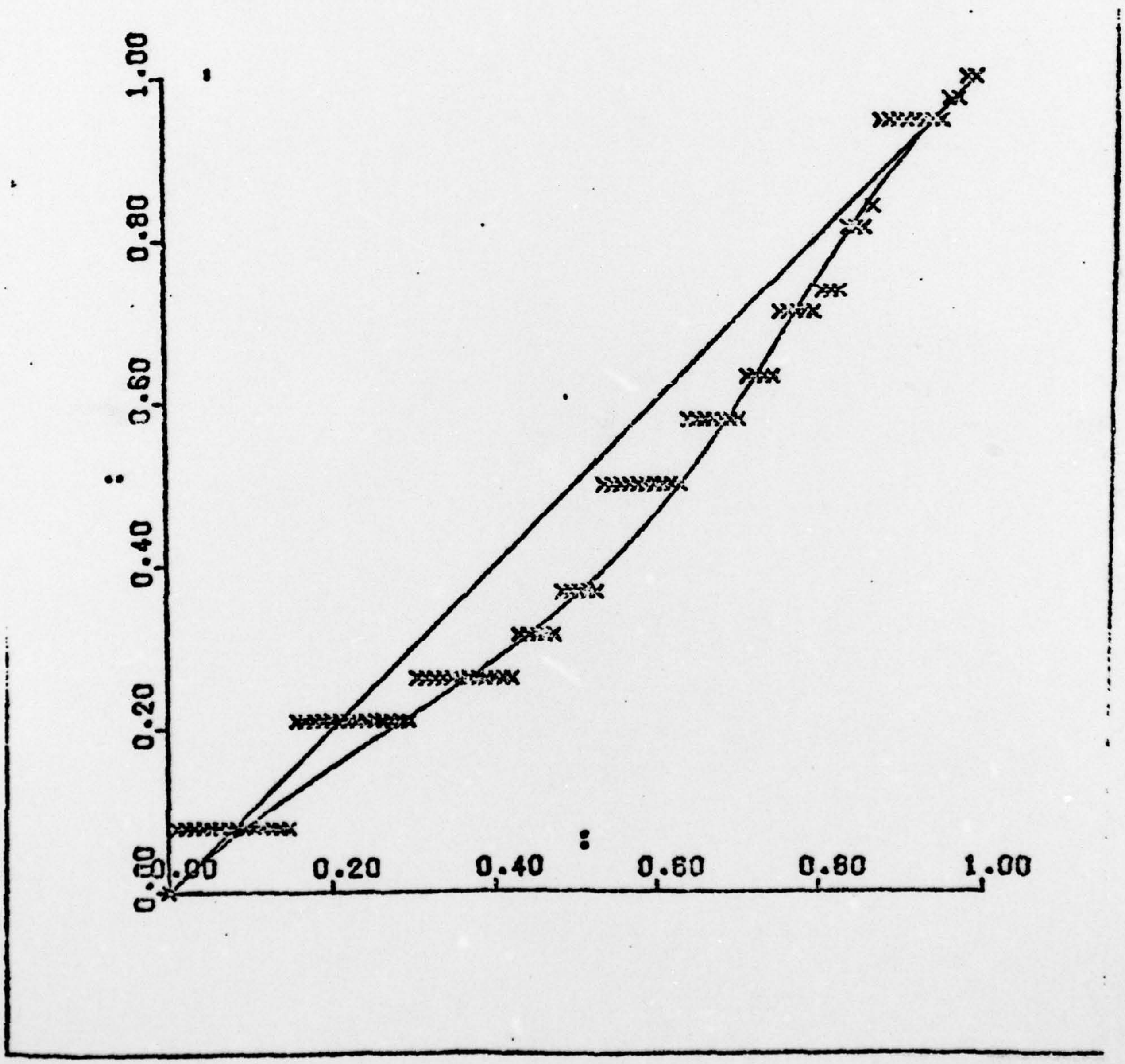


FIGURE 4.12

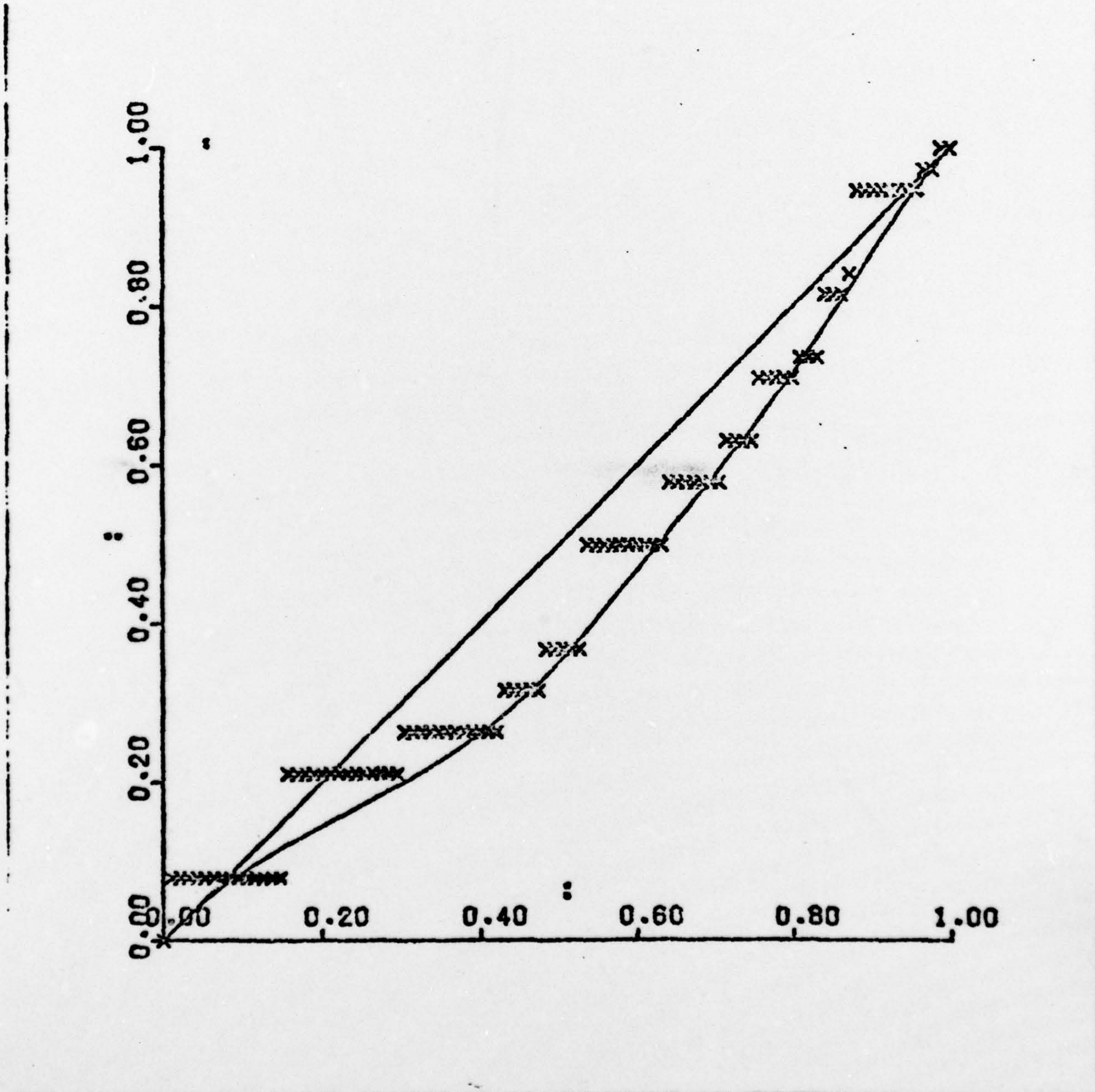


FIGURE 4.13

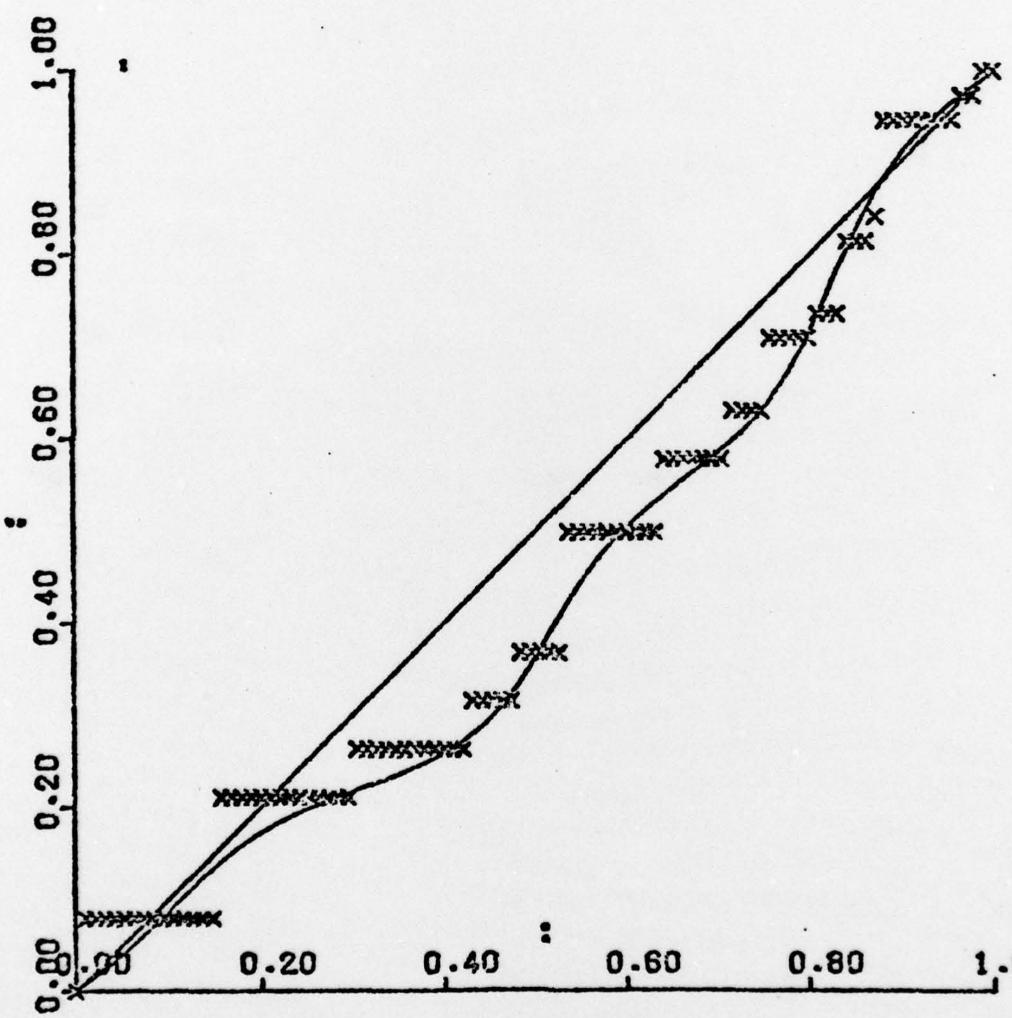


FIGURE 4.14

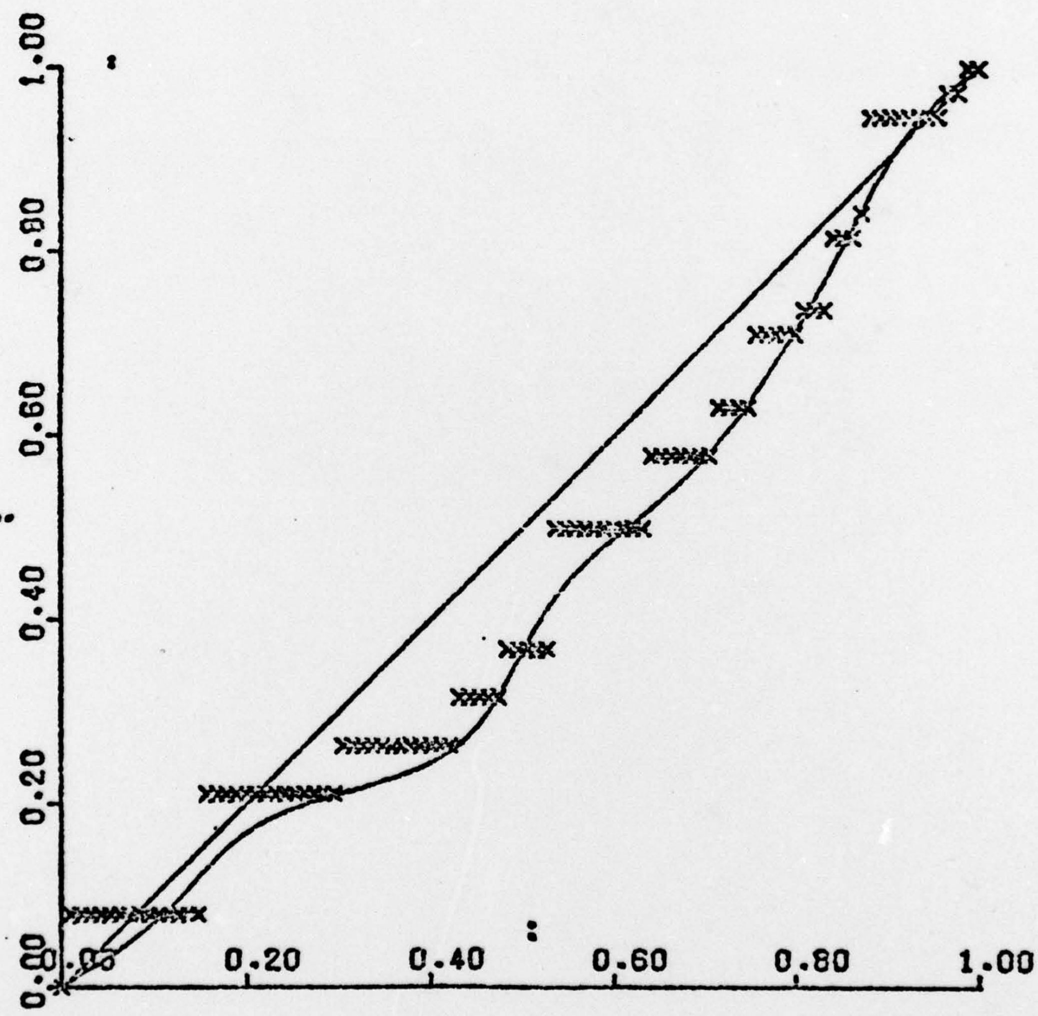


FIGURE 4.15

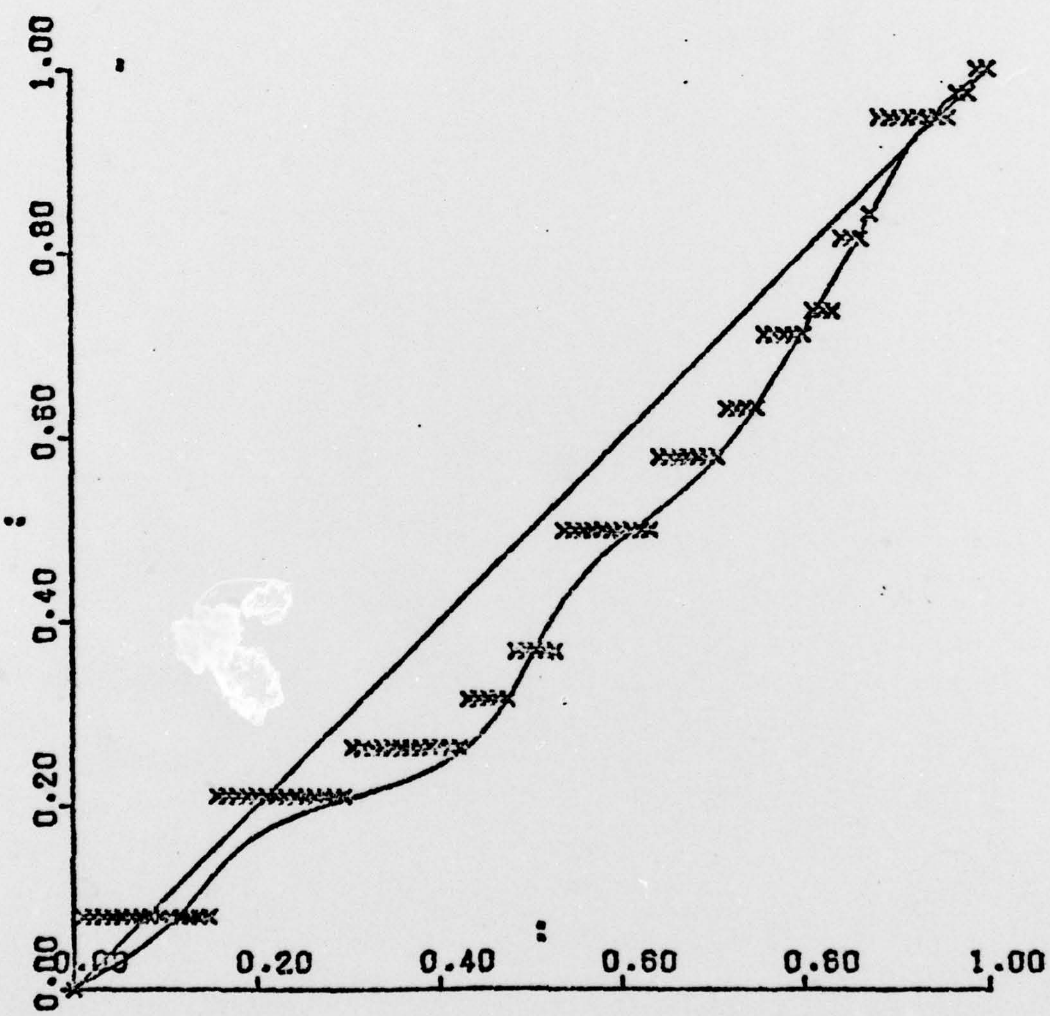


TABLE 4.7

Test	Observed	Critical Values (95%)
Wilcoxon	2203	1566 - 2082
Van der Waerden	12.69	-9.023 - 9.023
Median	29	13 - 24
Savage	47.33	28.87 - 47.13
$\lvert\tilde{\varphi}(1)\rvert^2$	.0492	.0316

Interpretations:

Even though the raw estimators of $\tilde{D}(\cdot)$ differ in appearance depending on whether we randomize the ties (Fig. 4.6-4.10) or allow for jumps of size greater than $1/N$ (Fig. 4.11-4.15), the smooth estimators are very comparable. In this case, they are below the line $D(u) = u$, meaning that the second group is smaller than the first one.

All the tests agree.

REFERENCES

Carmichael, J. P. (1976), The autoregressive method: a method of approximating and estimating positive functions, Technical Report No. 45, Statistical Science Division, SUNY at Buffalo.

Maguire, B. A. et al (1952), The time intervals between industrial accidents, Biometrika, 39, p. 168-180.

Parzen, E. (1977), Nonparametric statistical data science: a unified approach based on density estimation and testing for "white noise," Technical Report No. 47, Statistical Science Division, SUNY at Buffalo.

Welch, B. (1949), Further notes on Mrs. Aspin's tables, Biometrika, 36, p. 243-246.